

WORKING PAPER

VROEGE TRACKING, KWALITEIT EN RECHTVAARDIGHEID

Wat het wetenschappelijk onderzoek
ons leert over de hervorming van het
secundair onderwijs

Jeroen Lavrijsen, Ides Nicaise & Thomas Wouters

november 2013

KU LEUVEN

HIVA

ONDERZOEKSINSTITUUT VOOR ARBEID EN SAMENLEVING

VROEGE TRACKING, KWALITEIT EN RECHTVAARDIGHEID

Wat het wetenschappelijk onderzoek ons leert over de hervorming van het secundair onderwijs

Jeroen Lavrijsen, Ides Nicaise & Thomas Wouters

Abstract

De hervorming van het secundair onderwijs (Masterplan van 4 juni 2013) wil de oriënterende functie van de eerste graad versterken, zodat leerlingen pas op latere leeftijd een keuze moeten maken voor een bepaalde onderwijsvorm. Wetenschappelijk onderzoek, op basis van objectieve empirische data, toont aan dat zo'n latere oriëntering de impact van sociale afkomst op prestaties vermindert en dus de *rechtvaardigheid* van het onderwijs ten goede komt. Daarnaast suggereert de wetenschappelijke evidentie dat een latere oriëntering niet ten koste hoeft te gaan van de algemene *kwaliteit* van het onderwijs.

Dit rapport vat de omvangrijke wetenschappelijke literatuur hierover samen. Daarbij legt het ook uit welke belangrijke mankementen een eerder rapport van Duyck & Anseel (2012) hierover, waarin net het tegenovergestelde werd beweerd, vertoonde.

COMMENTAAR IS WELKOM jeroen.lavrijsen@kuleuven.be

KU Leuven
HIVA - ONDERZOEKSINSTITUUT VOOR ARBEID EN SAMENLEVING
Parkstraat 47 bus 5300, 3000 LEUVEN, België
hiva@kuleuven.be
www.hiva.be

© 2013 HIVA-KU Leuven

Niets uit deze uitgave mag worden veelelvuldigd en/of openbaar gemaakt door middel van druk, fotocopie, microfilm of op welke andere wijze ook, zonder voorafgaande schriftelijke toestemming van de auteur.

Inhoud

Inleiding	5
1 Onderwijs onderzocht	7
1.1 Essentiële begrippen	7
1.2 Mechanismes	9
1.3 Een eerste empirische toets	10
1.4 Scherpstellen naar een zuiverder beeld	11
2 Vroege tracking en rechtvaardigheid (sociale effecten)	13
2.1 De bewering van Duyck en Anseel	13
2.2 Wat het onderwijskundig onderzoek ons leert	13
2.3 Conclusie	17
3 Vroege tracking en kwaliteit (cognitieve effecten)	18
3.1 De bewering van Duyck en Anseel	18
3.2 Wat het onderwijskundig onderzoek ons leert	19
3.2.1 Landenvergelijkend onderzoek	19
3.2.2 Niet elke vorm van differentiëren is <i>tracking</i>	25
3.3 Conclusie	27
4 Algemene conclusie en aanbevelingen	28
5 Referenties	29

Inleiding

In juni 2013 kwam de Vlaamse Regering tot een akkoord over de hervorming van het secundair onderwijs. Een van de doelen van die hervorming was de oriënterende functie van de eerste graad te versterken, waardoor leerlingen pas op latere leeftijd zouden moeten kiezen voor een bepaalde onderwijsvorm (doorstromings- of arbeidsmarktgericht).¹

Welke gevolgen mogen we verwachten van zo'n latere oriëntering? Zal ons onderwijs er sociaal rechtvaardiger door worden? En gaat dat dan niet ten koste van de kwaliteit van ons onderwijs?

Het is belangrijk dat de antwoorden op dat soort vragen gebaseerd worden op wetenschappelijke gegevens, en niet louter op een ideologisch buikgevoel. Dat is precies de reden waarom organisaties zoals de OESO regelmatig internationaal vergelijkende scholientests laten uitvoeren, waarbij wordt nagegaan hoe leerlingen in verschillende landen scoren op wiskunde, leesvaardigheid of wetenschappen (het PISA-onderzoek). Op die manier kunnen verschillen tussen landen worden gedetecteerd, en kan worden nagegaan waarom sommige landen beter scoren dan andere.

Het wetenschappelijk onderzoek op basis van die empirische gegevens heeft geleid tot twee belangrijke conclusies.

De eerste is dat het vroeg opsplitsen van leerlingen *niet* leidt tot betere prestaties (*cognitieve effecten*). Er is dan ook geen enkele wetenschappelijk onderbouwde reden om aan te nemen dat het verbreden van de eerste graad zou leiden tot minder goede prestaties.

De tweede is dat het vroeg oriënteren van leerlingen nefast is voor de rechtvaardigheid van het onderwijs (*sociale effecten*). Hoe vroeger leerlingen worden opgedeeld, hoe sterker de invloed van sociale achtergrond op leerprestaties. Het verbreden van de eerste graad kan dan ook de rechtvaardigheid van het onderwijs vergroten.

Op basis van dit empirisch bewijsmateriaal spoort de OESO haar lidstaten dan ook met klem aan om leerlingengroepen niet te vroeg op te delen naar onderwijsvorm (OECD (2012): *postpone early tracking*). De geplande hervorming van het secundair onderwijs ligt in lijn met die internationale aanbeveling.

Toch lieten Wouter Duyck en Frederik Anseel vorig jaar plots een heel ander geluid horen. In hun discussiepaper voor de denktank Itinera (Duyck & Anseel (2012)) beweerden ze dat een vroege opsplitsing wél noodzakelijk zou zijn om tot kwaliteitsvol onderwijs te komen (*cognitief effect*). Bovendien, zo claimden ze, zou er geen bewijs zijn voor de stelling dat de vroege opsplitsing van leerlingen de impact van sociale afkomst op leerprestaties versterkt (*sociaal effect*).

¹ Daarnaast bevat het Masterplan een aantal andere maatregelen, bv. m.b.t. de kennis van het Nederlands (taalscreening, taalbad) of de plaats van techniek in het curriculum (vanaf het basisonderwijs). In deze paper beperken we ons echter tot de effecten van een latere oriëntering.

Goed onderwijs is voor onze maatschappij van levensbelang. Het is dan ook cruciaal dat ideeën over onderwijs worden onderbouwd door degelijk empirisch onderzoek. Het rapport van Duyck en Anseel heeft de verdienste dat het de onderwijshervorming heeft proberen te beoordelen op basis van wetenschappelijke evidentie, en niet op basis van ideologische vooronderstellingen. Het is dan ook spijtig dat hun rapport een aantal wetenschappelijke gebreken vertoont. Duyck en Anseel laten het grootste deel van het wetenschappelijk onderzoek over de effecten van het vroeg opsplitsen van leerlingen onvermeld, terwijl ander onderzoek verkeerd wordt voorgesteld.

In dit rapport proberen we de wetenschap de plaats in het onderwijsdebat te geven die ze verdient. Daarom gaan we dieper in op de wetenschappelijke evidentie die door Duyck en Anseel helaas onvermeld werd gelaten. We laten zien dat een uitgebreider overzicht van de wetenschappelijke literatuur suggereert dat een bredere eerste graad géén nadelige effecten hoeft te hebben voor de cognitieve prestaties, terwijl ze wenselijk is vanuit sociaal oogpunt.

1 | Onderwijs onderzocht

Op één punt zijn we het volmondig eens met Duyck en Anseel: het is belangrijk dat de onderwijshervorming beoordeeld wordt op basis van wetenschappelijk onderbouwde stellingen en *niet* op basis van ideologisch geïnspireerde opinies (Duyck & Anseel (2012), p. 2). Het betrekken van wetenschappelijk onderzoek in het maatschappelijke debat veronderstelt dat we de essentiële begrippen uit het wetenschappelijke jargon helder definiëren. Om het debat te faciliteren gebruiken we daarbij zoveel mogelijk dezelfde terminologie als Duyck en Anseel.

Daarna bekijken we waarom het vroegtijdig opsplitsen van leerlingen in aparte onderwijsvormen *in principe* zowel positieve als negatieve effecten zou kunnen hebben, en concluderen we dat enkel empirisch onderzoek, gebaseerd op objectieve data, ons iets kan leren over het *werkelijke* effect. Dat empirisch onderzoek zelf zal meer uitgebreid aan bod komen in hoofdstukken 2 en 3.

1.1 Essentiële begrippen

Onderwijs genereert onderwijsuitkomsten: leerlingen leren vaardigheden en kennis, verwerven een diploma en maken daar later op de arbeidsmarkt gebruik van.

Met de *kwaliteit* (*quality* of *efficiency*) van het onderwijs doelen we op het algemene niveau van die onderwijsuitkomsten. Die kwaliteit kan bijvoorbeeld worden afgelezen aan de prestaties van landen op vergelijkende scholientests zoals PISA. Duyck en Anseel hebben het in hun rapport over de “cognitieve effecten” van onderwijs.

De *rechtvaardigheid* van het onderwijs (*equity* of *equality of opportunity*) duidt de mate aan waarin iemands sociale afkomst zijn onderwijsuitkomsten bepaalt. Een zeker effect van SES (socio-economische status) op prestaties is onvermijdelijk, omdat intelligentie voor een deel genetisch wordt bepaald en dus niet evenredig verdeeld hoeft te zijn over sociale klassen (cf. Fischbein, 1980). Ongeacht het onderwijssysteem zal sociale afkomst dus voor een deel de prestaties beïnvloeden. Waar het hier om gaat, is dat de band tussen afkomst en prestaties in sommige landen kleiner is dan in andere. Hoe kleiner het effect van sociale afkomst op prestaties, hoe rechtvaardiger we het onderwijs kunnen noemen. Duyck en Anseel noemen dat de sociale effecten van het onderwijs.

We doelen met rechtvaardigheid (“gelijke kansen”) dus expliciet *niet* op “gelijke uitkomsten” (kleine verschillen tussen sterke en zwakke leerlingen). Daar waar een evolutie in de richting van meer rechtvaardigheid altijd positief is, kunnen “gelijke uitkomsten” niet zo eenduidig worden gewaardeerd. Een evolutie waarbij het niveau van de sterkste presteerders afneemt, terwijl dat van de zwakkeren gelijk blijft, is bijvoorbeeld geenszins positief.

Tracking is het opsplitsen van leerlingengroepen in verschillende onderwijsvormen (in Vlaanderen in aso/tso/bso/kso) met elk een specifieke invulling en een uitgesproken finaliteit (hoger onderwijs, arbeidsmarkt). De verschillende onderwijsvormen worden maatschappelijk niet gelijk gewaardeerd, waarbij theoretische, abstracte *tracks* hoger gewaardeerd worden. Dat is ook zichtbaar in het woordgebruik, waarbij er over “hogere” en “lagere” onderwijsvormen wordt gesproken². De opsplitsing is definitief van aard: veranderen kan doorgaans enkel nog “naar beneden”. Bovendien worden de verschillende vormen doorgaans aangeboden in aparte scholen.

Elk land in Europa maakt vroeg of laat gebruik van *tracking*. Het punt waar het in deze discussie om gaat, is dus *niet* dat leerlingen die van elkaar verschillen in talenten en interesses op een zeker moment kiezen voor een andere onderwijsvorm. Waar het wél om gaat, is *de leeftijd waarop* dit gebeurt. Op dit punt zijn er grote verschillen tussen landen: “*vroege trackers*” zoals Duitsland splitsen hun leerlingen al op op 10 jaar, terwijl “*comprehensieve landen*” zoals Finland ermee wachten tot 16 jaar. Merk daarbij op dat men ook in comprehensieve landen rekening houdt met verschillen tussen leerlingen (bv. door te voorzien in extra remediërende en verdiepende cursussen voor wie daar behoefte aan heeft); het onderscheidend punt tussen beide systemen is dus of het leidt tot een definitieve opdeling over de hele lijn (zoals bij *tracking*) of niet. We komen daar later nog op terug.

Officieel kent de Vlaamse eenheidsstructuur pas een opsplitsing op 14 jaar. De school- en studiekeuze op 12 jaar is echter sterk bepalend voor het vervolg van de onderwijsloopbaan, enerzijds omdat veel secundaire scholen in hun bovenbouw slechts één onderwijsvorm aanbieden en anderzijds omdat al in de eerste graad de keuze voor bepaalde optievakken (Latijn, technologie ...) de latere studiemogelijkheden sterk beperkt. *De facto* gebeurt de opsplitsing in Vlaanderen dus al op 12 jaar. Dat plaatst ons dus eerder in de groep van de landen met *vroege tracking*.

Het Masterplan Hervorming S.O. van 4 juni 2013 voorziet nu precies een aanpassing van de structuur van het secundair onderwijs, met als expliciet doel de “*oriënterende functie van de eerste graad te versterken*” en de definitieve studieoriëntering uit te stellen tot 14 jaar. Enerzijds gebeurt dit door het aanmoedigen van domeinscholen (die niet gericht zijn op één onderwijsvorm maar op het studiedomein) en anderzijds door de bestaande optiepakketten in de eerste graad anders te gaan invullen (bv. door meer aandacht te hebben voor hun remediërende functie)³. Langs de andere kant laat het Masterplan - als politiek compromis - nog veel details open, waardoor het nog niet duidelijk is hoe het in de praktijk precies zal worden uitgewerkt. In dit rapport focussen we dan ook niet zozeer op de concrete invulling van het Masterplan, maar wel op de meer fundamentele vraag of zo een uitstel van oriëntering überhaupt wel een goed idee is. In de wetenschappelijke literatuur zullen we dus op zoek moeten gaan naar een antwoord op de vraag: wat zijn de positieve en negatieve gevolgen van *early tracking*?

2 Omdat die bewoordingen sterk in het onderwijsdiscours ingeburgerd zijn, nemen we de terminologie in deze paper over, zonder daarmee een oordeel uit te spreken over de waarde van de verschillende onderwijsvormen.

3 De overgang naar domeinscholen wordt aangemoedigd maar niet verplicht. Wanneer scholen zich echter massaal zouden blijven profileren in functie van hun bovenbouw (arbeidsmarkt- of doorstromingsgericht), zal de eerste graad in de praktijk mogelijk even smal blijven als vandaag. Een gelijkaardig voorbehoud maken we bij de invulling van de optiepakketten: indien die een even determinerende voorafname op de latere studiekeuze blijven als vandaag, zal deze hervorming de eerste graad niet echt breder maken.

1.2 Mechanismes

Vroege tracking kan de leerprestaties op verschillende manieren beïnvloeden.

Specialisatie-effect

Een voordeel van *tracking* is dat het klasgroepen homogener maakt. Als de leerlingen in de klas ongeveer gelijke capaciteiten hebben, dan kan het tempo, de leerstof en de instructiemethode eenvoudiger op dat niveau worden afgestemd. Een vroege opdeling kan dan zowel zwakke als sterke leerlingen ten goede komen (Brunello & Checchi (2007)).

Misallocatie-effect

Het bovenstaande argument veronderstelt dat de keuze voor een bepaalde onderwijsvorm volledig bepaald wordt door de capaciteiten van de leerling. In werkelijkheid worden studiekeuzes echter óók bepaald door sociale afkomst, los van het effect van vroegere schoolresultaten (Boone & Van Houtte (2012)). Wanneer sterke maar kansarme leerlingen in minder sterke tracks terechtkomen puur op basis van hun sociale achtergrond, dan gaat het specialisatie-effect uiteraard niet op. Internationaal onderzoek (cf. Brunello & Checchi (2007)) suggereert dat de versturende impact van sociale achtergrond groter wordt naarmate de *tracking* eerder gebeurt (bv. omdat rijpere leerlingen meer los kunnen komen van sociale verwachtingen). *Vroege tracking* kan op die manier dus ook leiden tot een verspilling van talent.

Peer- en omgevingseffecten

Ook de sterkte van de klasgroep beïnvloedt de individuele leerprestaties. Sommige onderzoekers beweren daarbij dat zwakke leerlingen gebaat zijn bij de aanwezigheid van sterkere klasgenoten, terwijl sterke leerlingen weinig gevoelig zijn voor de invloed van (zwakkere of even sterke) leerlingen. Er bestaan echter verschillende ideeën over de sterkte en de richting van die effecten.

Wellicht belangrijker is dan ook het feit dat veel leerlingen in de “lagere” tracks daar terecht zijn gekomen door een negatieve keuze (omdat men niet beantwoordde aan de vereisten voor een hogere track). In Vlaanderen toonde Van Houtte (2006) aan dat een aantal leerlingen deze “faalervaring” verwerken door op zoek te gaan naar alternatieve bronnen van status (bv. wangedrag). In tso- en bso-scholen is de cultuur daardoor veel minder gericht op studeren, óók als gecorrigeerd wordt voor evidente verschillen in capaciteiten. Het rigide karakter van de *tracks* – waaruit niet meer kan worden “opgeklommen” – zorgt dan weer voor wijdverspreide gevoelens van zinloosheid (futiliteit, Van Houtte & Stevens (2010)). Van Houtte (2004) toont aan dat ook leerkrachten hun verwachtingspatronen naar beneden stellen (opnieuw onafhankelijk van de werkelijke capaciteiten van de leerlingen). Alles samen leidt dit ertoe dat leerlingen die in de basisschool gelijke cijfers haalden, een stuk minder goed gaan presteren wanneer ze in een tso- of bso-school terecht komen in plaats van in een aso-school (Van Houtte (2004)). Dit bedreigt de leerkansen van de leerlingen in kwestie en de kwaliteit van het technisch- en beroepsonderwijs als geheel.

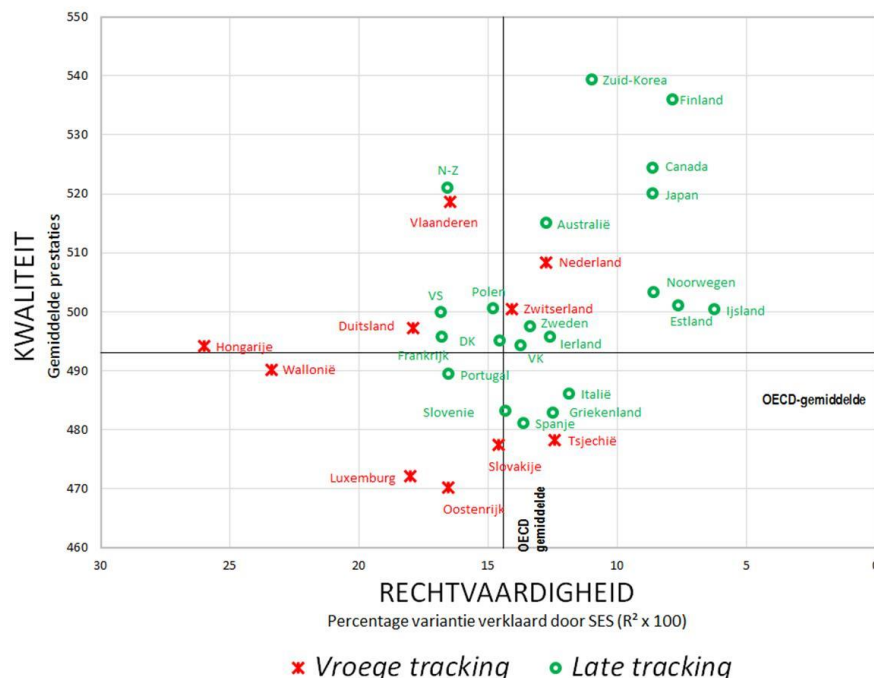
Maatwerk kan ook zonder tracking

Een laatste belangrijke kanttekening bij de specialisatie-hypothese is dat ze onderschat dat ook een comprehensief systeem leerlingen met verschillende capaciteiten op een aangepaste manier kan begeleiden. Een goed voorbeeld is de manier waarop Finland (*tracking* pas op 16 jaar) omgaat met leerlingen die moeilijkheden ondervinden voor een bepaald vak. Die leerlingen worden voor het vak in kwestie een tijdje apart genomen en intensief bijgewerkt, terwijl de rest van de groep de tijd benut om de basisleerstof te verdiepen. Nadat het tekort is weggewerkt, sluiten de leerlingen gewoon weer aan bij de rest van de klasgroep (met wie ze voor de andere vakken samen zijn

gebleven). Die aanpak leidt dus óók tot een specialisatievoordeel, maar is volstrekt niet gelijk te schakelen met *tracking*: de opsplitsing is slechts tijdelijk (in plaats van de definitieve plaatsing in een onderwijsvorm die de rest van de carrière sterk bepaalt), de remediëring is expliciet gericht op een inhaalbeweging waardoor zwakkere leerlingen terug kunnen aanpakken bij het basisniveau (i.p.v. het creëren van minder veeleisende tracks), en de opdeling gebeurt enkel voor die vakken waar de leerling moeilijkheden voor ondervindt (i.p.v. een opdeling voor het volledige curriculum). Finland haalt met deze aanpak uitmuntende resultaten op internationaal vergelijkende tests.

1.3 Een eerste empirische toets

De hierboven beschreven theoretische kanalen gaan dus verschillende richtingen uit: sommige pleiten in het voordeel van *vroege tracking*, andere in het nadeel. Om vast te stellen of vroege tracking nu een voor- of een nadeel is, zullen we dus empirisch moeten gaan kijken naar de daadwerkelijke prestaties van de verschillende onderwijssystemen.



Figuur 1: Kwaliteit en rechtvaardigheid in PISA2009 (leesvaardigheid 15-jarigen)

Deze empirische gegevens vinden we met name in PISA, de internationaal gestandaardiseerde test waarmee om de drie jaar de wiskundige, wetenschappelijke en leesvaardigheden van 15-jarigen in verschillende landen worden gemeten. Ook de TIMSS-peilingen, die worden uitgevoerd in het 2^e jaar van het secundair onderwijs, geven informatie over de leerprestaties in landen. Een belangrijk voordeel van PISA is dat de geteste scholieren al 1 à 2 jaar ouder zijn dan die in TIMSS. De scholieren in PISA hebben dus al langer de effecten van *tracking* ondervonden, waardoor die beter zichtbaar zullen zijn.

Wat leert PISA ons nu over het effect van *vroege tracking*? Figuur 1 geeft hierop een eerste empirisch antwoord. De verticale as bevat de landenscores voor lezen in PISA (jaargang 2009). De landen die het hoogst staan, presteren het best. De horizontale as geeft de impact van sociale afkomst aan. In

de landen aan de rechterkant bepaalt SES de leerprestaties het minst: deze zijn dus het rechtvaardigst. De landen in de rechterbovenhoek combineren dus een hoge kwaliteit met een hoge rechtvaardigheid. Landen waarin leerlingen vroeg getracked worden (ten laatste op hun 13^{de}), worden aangeduid met een rood kruisje, terwijl landen waarin de tracking langer wordt uitgesteld worden aangeduid met een groen bolletje.

Op het eerste zicht suggereert de figuur dat *vroege tracking* absoluut niet nodig is om tot kwaliteitsvol onderwijs te komen. Meer zelfs: de kopgroep van best presterende landen bestaat vooral uit onderwijssystemen waar de opdeling pas laat wordt gemaakt. Anderzijds suggereert de figuur dat *vroege tracking* nadelig is voor de *rechtvaardigheid* van het onderwijssysteem. De vroege trackers bevinden zich in de regel meer aan de linkerkant van de grafiek, terwijl aan de rechterkant vooral comprehensieve landen voorkomen. Een vroege opdeling lijkt dus het effect van sociale afkomst op de leerprestaties te versterken.

Vlaanderen bevindt zich op Figuur 1 in de linkerbovenhoek. Dit betekent ten eerste - en dat mag zeker niet vergeten worden - dat onze leerlingen gemiddeld erg goede prestaties halen. Tegelijkertijd is de band tussen sociale afkomst en prestaties bij ons sterker dan gemiddeld. Een aantal andere landen slaagt er blijkbaar een stuk beter in om goede algemene prestaties te combineren met gelijke kansen.

De uitdaging voor het Vlaamse onderwijs bestaat er dan ook in om van linksboven naar rechtsboven te evolueren: de goede prestaties behouden, maar in een rechtvaardiger systeem. De figuur suggereert dat een uitgestelde studiekeuze inderdaad een manier is om dat te bewerkstelligen: late trackers scoren gemiddeld minstens even goed als vroege trackers, maar de band tussen afkomst en prestaties wordt over algemeen minder sterk als de studieoriëntering wordt uitgesteld.

1.4 Scherpstellen naar een zuiverder beeld

Toch zijn de bewijsvoeringen uit de vorige paragraaf nog te eenvoudig. Een *correlatie* tussen twee variabelen (leeftijd van *tracking* enerzijds en onderwijsuitkomsten anderzijds) betekent immers nog niet dat er ook een *oorzakelijk verband* bestaat. Onderwijsuitkomsten worden immers ook door heel wat andere factoren beïnvloed. Een voorbeeld: welvarende landen scoren doorgaans beter dan armere landen. Zijn de goede prestaties van de landen rechtsboven dan misschien gewoon een gevolg van het feit dat het welvarende landen zijn? Zijn de verbanden tussen *vroege tracking* en onderwijsuitkomsten dus wel werkelijk bestaande verbanden, of liggen er onrechtstreeks andere “verstorende” variabelen aan ten grondslag?

In de literatuur wordt dit het “*omitted variables*”-probleem genoemd: als in een model bepaalde verklarende variabelen (zoals bijvoorbeeld welvaart) niet opgenomen zijn, dan kan dit het effect van andere variabelen verstoren. Gelukkig bestaan er verschillende statistische technieken om die valkuil te omzeilen.

Controlevariabelen

De eerste optie bestaat erin om de verstorende variabelen zelf mee in de modellen op te nemen en zo “onder controle te houden”. Door van de beschouwde landen het BBP/capita in rekening te brengen, kunnen we het verstorende effect van welvaartsverschillen uitzuiveren en een beter beeld krijgen van het effect van *tracking* zelf. Een probleem daarbij is dat we nooit helemaal zeker kunnen

weten of we alle relevante verschillen tussen landen wel hebben meegenomen. Dat betekent ook dat we een voorkeur hebben voor relatief homogene datasets: het verstorende effect van niet-geobserveerde verschillen is uiteraard kleiner wanneer de onderzochte landen voldoende op elkaar lijken dan wanneer het om volstrekt verschillende landen gaat. Meestal focust men dan ook op een subset van landen met een vergelijkbaar ontwikkelingsniveau, bv. de leden van de OESO.

Differences-in-differences-methode

Een tweede oplossing is om te controleren voor verschillen in de “startpositie” van verschillende landen. We hebben immers niet alleen informatie over de prestaties van de 15-jarigen in de verschillende landen (PISA), maar we weten ook hoe die landen het deden in het basisonderwijs (TIMSS bevat immers ook een module voor het 4^e leerjaar). Als welvaart effectief de onderwijsprestaties beïnvloedt, dan doet het dat voor beide tests. Op de *toename* van de vaardigheden tussen de twee meetpunten heeft welvaart (of eender welke andere verstorende variabele) dus geen effect meer. Het effect van *tracking* kan dan zuiverder worden afgelezen uit de verschillen tussen landen in de toename tussen de twee meetpunten (vandaar: *differences-in-differences*).

Experimentele situaties

Een laatste mogelijkheid bestaat erin om het effect van een onderwijshervorming in één bepaald land te bekijken. De voorbije decennia hebben heel wat landen de leeftijd van *tracking* opgeschoven. Door de prestaties voor en na de hervorming te vergelijken, kunnen we quasi-experimenteel vaststellen wat het effect ervan is geweest.

Elk van de drie hierboven geschetste technieken kent zijn voor- en zijn nadelen. Stellingen krijgen dan ook meer wetenschappelijke gewicht naarmate ze worden onderbouwd door studies uit elk van die onderzoekstechnieken.

2 | Vroege tracking en rechtvaardigheid (sociale effecten)

In dit hoofdstuk bekijken we wat het internationaal vergelijkend onderzoek ons leert over de invloed van *vroege tracking* op de rechtvaardigheid van het onderwijs. Zoals hoger aangegeven, bedoelen we hiermee de mate waarin de sociale afkomst leerprestaties beïnvloedt.

2.1 De bewering van Duyck en Anseel

Zoals we zullen zien is de bespreking van de sociale effecten van *vroege tracking* wellicht het merkwaardigste stuk van het rapport van Duyck en Anseel.

We citeren (p. 10):

*“Om de invloed van vroege tracking op SES-effecten (...) te kunnen bepalen is per definitie internationale vergelijking nodig, zodat de Vlaamse SES effecten kunnen vergeleken worden met SES effecten in systemen met late tracking. Zijn de SES effecten daar kleiner? Het wetenschappelijke antwoord op deze vraag is teleurstellend. Er zijn veel studies die effecten van SES op leerprestaties bestuderen (vaak in sociologische of economische tijdschriften), en studies die effecten van vroege tracking op leerprestaties bestuderen. **Het blijkt echter veel minder evident in de literatuur een antwoord te krijgen op de vraag of vroege tracking de effecten van SES op leerprestaties beïnvloedt** (in wetenschappelijke terminologie: of vroege tracking met SES interageert): zelfs al bestaan er sociale effecten op studiekeuze en leerprestaties, worden deze veroorzaakt of tenminste beïnvloed door vroege tracking? In geen enkele van de sociologisch geïnspireerde opiniestukken in het debat wordt expliciet naar een dergelijk effect in een gepubliceerd wetenschappelijk artikel verwezen. (...) Het blijft onduidelijk of vroege tracking leidt tot grotere sociale discriminatie effecten dan late tracking.”*

2.2 Wat het onderwijskundig onderzoek ons leert

Duyck en Anseel beweren dus dat de invloed van *vroege tracking* op de rechtvaardigheid van het onderwijs nog onvoldoende onderzocht zou zijn - en bespreken dan ook geen enkel onderzoek over deze kwestie. Dit is een nogal eigenaardige bewering. De impact van *vroege tracking* op het SES-effect – het interactie-effect waar Duyck en Anseel naar op zoek zijn – is al wel degelijk het voorwerp geweest van heel wat wetenschappelijk onderzoek. De literatuur hierover is inmiddels al zo uitgebreid geworden dat overzichtswerken (reviews) werden opgesteld om de veelheid aan studies samen te vatten.

Een goede synthese verscheen bijvoorbeeld enkele jaren geleden in het gereputeerde wetenschappelijke tijdschrift *Annual Review of Sociology* (Van de Werfhorst & Mijs (2010)). Daarin werd op basis van een “*vast and growing number of studies*” beargumenteerd dat *early tracking* het effect van SES op de prestaties onmiskenbaar versterkt:

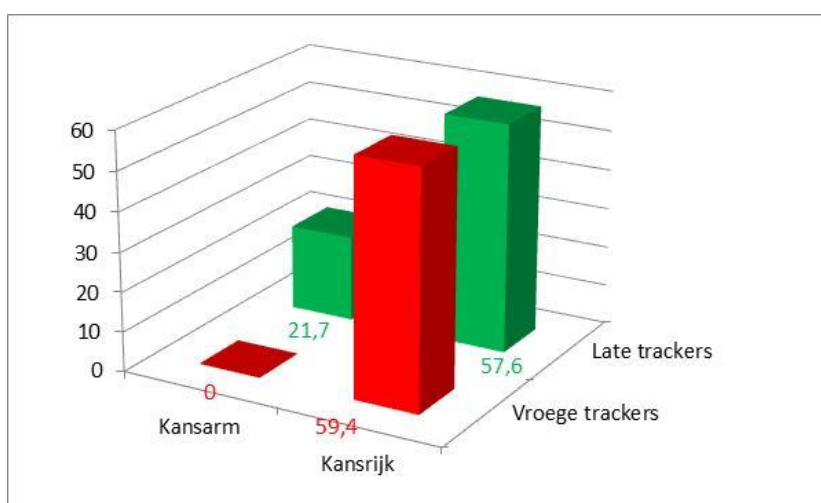
“the dependence of achievement on social class [is] higher in educational systems that track students in different school types and school locations relative to systems that offer comprehensive schools. This effect is sizeable and is confirmed by national studies on the mechanisms underlying tracking effects. Moreover [...] this finding is disputed rarely.”

Uit deze objectieve review van de wetenschappelijke literatuur komt dus een sterke consensus naar voren: *vroege tracking* versterkt het SES-effect. Andere reviews (bv. Hanushek & Woessmann (2010)) komen trouwens tot precies dezelfde conclusie.

We zullen hieronder kort ingaan op een aantal individuele studies die de conclusie uit de bovenstaande literatuurreviews onderschrijven, waarbij we onderzoeken uit elk van de bovenstaande onderzoeksdesigns (zie 1.4) bekijken. Het feit dat studies met verschillende designs en met verschillende datasets de stelling onderbouwen, vergroot immers de geldigheid van die conclusie.

1. Een eerste illustratie van de impact van *vroege tracking* op het SES-effect is de analyse van Woessmann, Luedemann, Schuetz & West (2009). Op basis van PISA2003 laten ze zien hoe SES een belangrijke voorspeller is van onderwijsprestaties: gemiddeld genomen presteren kinderen met een hoge SES beter. Zoals hoger aangehaald, is dit wellicht voor een deel het onvermijdelijke gevolg van het feit dat intelligentie niet uniform verdeeld is over sociale klassen. Wat ons vooral interesseert, is hun vaststelling dat dit effect van land tot land varieert, waarbij het effect groter wordt naarmate een land leerlingen vroeger oriënteert. *Vroege tracking* vergroot dus inderdaad het effect van sociale afkomst op leerprestaties.

Figuur 2 vat de gevonden relatie samen. Hierbij worden de gemiddelde scores voor wiskunde uitgedrukt t.o.v. de referentiegroep van kinderen met een lage SES in vroeg trackende landen. Hun tegenhangers in laat trackende landen scoren gemiddeld meer dan 20 punten beter. Langs de andere kant maakt het voor kinderen met een hoge SES minder uit wanneer *getrackt* wordt: het verschil in punten is miniem. Dit impliceert twee zaken. Ten eerste is de impact van SES op prestaties (de kloof tussen kinderen met een lage en een hoge SES) veel groter in de vroeg *trackende* landen. Ten tweede zijn de gemiddelde prestaties over alle sociale klassen minder goed in de vroeg *trackende* landen. Op dit laatste komen we terug in hoofdstuk 3.



Figuur 2: Vroege tracking vergroot het effect van sociale afkomst op leerprestaties, overgenomen uit Woessmann, Luedemann, Schuetz & West (2009).

2. De stelling dat *vroege tracking* de impact van sociale afkomst op leerprestaties vergroot wordt ook bevestigd in de analyse van Horn (2009) voor het wetenschappelijke tijdschrift *Educational research and evaluation*. Daarbij wordt gebruik gemaakt van de statistisch meer gesofisticeerde techniek van de multilevelanalyse, wat een erg geschikte methode is om kenmerken op het leerlingenniveau (sociale afkomst) te interageren met kenmerken op het landniveau (de leeftijd van *tracking*). Controlerend voor een aantal andere kenmerken van het onderwijssysteem (bv. het gebruik van centrale examens en de schoolautonomie) wordt daarbij opnieuw ondubbelzinnig vastgesteld dat een vroege *tracking* de onrechtvaardigheid van het onderwijs vergroot.

“The most policy-relevant finding is that the early age of selection links closely with high inequality of opportunity. Age of selection associates very closely with the socioeconomic background effect, or in other words, the higher the countries select, the higher their inequality of opportunity is. The significance of the age of selection in increasing inequality of opportunity is unsurprising but very important.”

3. Ook het artikel van Dupriez, Dumay & Vause (2008) in het wetenschappelijke tijdschrift *Comparative Education Review* komt tot de conclusie dat een te vroege opdeling het effect van SES op leerprestaties vergroot. De band tussen sociale afkomst en PISA-prestaties is ook hier het grootst in vroeg trackende landen:

“Early tracking systems are characterized by the highest degree of social inequality”.

Ze geven daarbij aan dat hun conclusie niet op zichzelf staat, maar volledig in lijn ligt met wat de uitgebreide literatuur daarover al leerde:

“The stream of research has shown that school systems based on an early tracking process are characterized by a stronger relationship between students’ performance and sociocultural background”.

4. Onderzoek op basis van de meest recente jaargang van PISA (2009) onderbouwt verder de stelling dat *vroege tracking* het SES-effect versterkt. Zo tonen Bol & Van de Werfhorst (2013) in *Comparative Education Review* aan dat, controlerend voor welvaart, de hoogte van de uitgaven voor onderwijs en andere onderwijssysteemkenmerken, het effect van sociale afkomst groter is in systemen met *vroege tracking*:

“In more tracked educational systems, variation in student performance is more strongly based on social class background. Tracking enhances the importance of social origin for reading performance.”

5. Op basis van weer een andere PISA-jaargang (2000) laten ook Duru-Bellat & Suchaut (2005) in *European Educational Research Journal* zien dat landen verschillen in de mate waarin sociale afkomst de leerprestaties bepaalt, en dat het met name *vroege tracking* is die deze variatie in de grootte van deze sociale effecten verklaart:

“[early tracking] tends to increase social inequality in performance”

6. Ook wanneer andere scholientests worden gebruikt, blijft de boodschap gelijklopend. Zo laten Schütz, Ursprung & Woessmann (2008) zien dat ook op basis van TIMSS moet worden geconcludeerd dat het effect van sociale afkomst op leerprestaties toeneemt als leerlingen vroeger worden opgesplitst in verschillende onderwijsvormen:

“the family-background effect is larger, the earlier a country tracks its students into different school types.”

7. Dupriez & Dumay (2006) gaan meer in detail na of de correlatie tussen *vroege tracking* en onrechtvaardigheid niet verklaard kan worden door de impact van andere contextvariabelen, zoals maatschappelijke ongelijkheid. Hun conclusie is dat dat niet het geval is: het is wel degelijk de *vroege tracking* die zorgt voor de sterkere impact van SES op leerprestaties:

“The structure of the school system has a specific effect on the extent of inequalities. (...) School systems without tracking before age 16 tend to be able to reduce inequality of opportunities more than school systems with tracking from age 10, 11 and 12.”

8. De bovenstaande artikels controleren steeds voor een aantal verschillen tussen landen om de mogelijke impact van versturende variabelen uit te zuiveren (zie 1.4). Onze laatste drie voorbeelden hanteren allen een ander onderzoeksdesign om meer diepgaand na te gaan of de gevonden nefaste effecten van *vroege tracking* toch niet veroorzaakt werden door ongeobserveerde verschillen.

Een eerste slimme manier om hier een antwoord op te krijgen is de studie van federale staten, zoals Duitsland. De interne socio-economische en culturele verschillen tussen de Duitse deelstaten zijn immers in de regel minder groot dan de verschillen tussen landen. Woessmann (2010) laat zo zien dat ook tussen de Duitse deelstaten onderling *vroege tracking* een negatief effect heeft op kansengelijkheid: het effect van SES op leerprestaties wordt groter naarmate de deelstaat later *trackt*. Bovendien ligt de grootte van het effect in lijn met de effecten die in eerdere onderzoeken (op OESO-niveau) gevonden worden. Dit versterkt dus de betrouwbaarheid van de resultaten, en weerlegt de kritiek dat de in landenvergelijkende studies gevonden effecten van *vroege tracking* veroorzaakt worden door andere, niet-geobserveerde verschillen tussen die landen:

“The same results found previously across countries hold within Germany: (...) higher equality of opportunity is associated with reduced tracking.”

9. In 1.4 identificeerden we ook de “differences-in-differences”-benadering als een mogelijke techniek om na te gaan of de effecten van *vroege tracking* niet verklaard kunnen worden door andere verschillen tussen landen. Hierbij worden de PISA-resultaten (op 15 jaar) gecorrigeerd voor de “startpositie” van elk land in het basisonderwijs. Ammermüller (2005) bekijkt hoe de impact van SES op leerprestaties evolueert tussen het 4e leerjaar (PIRLS) en de leeftijd van 15 jaar (PISA). Hij stelt vast dat de impact van SES op leerprestaties veel sterker toeneemt in landen waar leerlingen sneller getrackt worden:

“greater differentiation of the schooling system as indicated by streaming (...) is associated with a greater effect of social background”.

10. Een laatste manier om het effect van *vroege tracking* zo zuiver mogelijk te bepalen, is te kijken naar experimentele situaties: hoe evolueert de band tussen sociale afkomst en prestaties wanneer in een land de leeftijd van *tracking* wordt uitgesteld?

Pekkarinen, Uusitalo, and Pekkala Kerr (2013) bekijken zo het effect van de comprehensivering van het Finse onderwijs in de jaren '70. Vóór die periode kende Finland een onderwijssysteem dat ongeveer vergelijkbaar was met het huidige Vlaamse systeem: leerlingen werden al vroeg naar een onderwijsvorm georiënteerd die de rest van hun onderwijs carrière bepaalde (algemeen vs. beroepsonderwijs). Ná de hervorming werd het Finse onderwijs comprehensief: leerlingen bleven tot hun 16de samen en maakten pas daarna een keuze voor een bepaalde track.

In vergelijking met landenvergelijkende studies zijn dit soort onderzoeken interessant omdat de mogelijk verstorende invloed van andere verschillen tussen landen (welvaart, cultuurverschillen ...) kan worden uitgesloten. Finse leerlingen worden immers vergeleken met Finse leerlingen; het enige verschil tussen beiden is de “behandeling” die ze kregen (sommigen werden vroeg getracked, anderen pas laat). De Finse hervorming is extra interessant omdat ze geleidelijk doorheen het land werd uitgerold. Dit betekent dat er tussen ’72 en ’78 delen van het land waren waarin het onderwijs al comprehensief was, terwijl in andere delen *vroege tracking* nog in zwang was. Hierdoor kan het effect van de onderwijs hervorming nog beter worden bestudeerd: op elk moment bestond er immers een groep leerlingen die al de gevolgen van de latere oriëntering ondervond, terwijl een “controlegroep” nog onder het *vroege tracking*-regime les kreeg.

Pekkarinen e.a. kwantificeren de effecten van de hervorming door te kijken naar de resultaten van cognitieve tests die werden afgenomen bij het begin van de verplichte legerdienst. Ze stellen vast dat:

“the reform significantly improved the scores in all tests for students whose parents had only basic education or low income. At the same time, the reform had no negative effects on the test scores of students from more advantaged backgrounds.”

Opnieuw blijkt hier dus uit dat de latere oriëntering het effect van sociale afkomst verkleint: kinderen met een lage SES beginnen het beter te doen en dichten zo een deel van de kloof met de kinderen met een hoge SES. Merk opnieuw op dat die laatste het niet minder goed beginnen te doen, en dat de gemiddelde kwaliteit dus zal toenemen – zie hoofdstuk 3.

Ook studies naar de effecten van hervormingen zélf bevestigen dus de duidelijke boodschap: *vroege tracking* versterkt de impact van SES op leerprestaties.

We hebben ons in dit overzicht gefocust op studies die de impact van *tracking* op de band tussen SES en *leerprestaties* onderzoeken, om zo dicht mogelijk bij de oorspronkelijke opzet van Duyck en Anseel te blijven. Het is in principe ook mogelijk om de band tussen sociale achtergrond en andere onderwijsuitkomsten te bestuderen. Brunello & Checchi (2007) tonen bijvoorbeeld al aan dat *early tracking* ook na de intrede op de arbeidsmarkt blijft doorwerken, en dat het bv. de impact van sociale afkomst op inkomen versterkt. Zelf lieten we zien (Lavrijsen & Nicaise, 2013) dat *early tracking* de impact van sociale afkomst op het halen van een secundair diploma vergroot: in vroeg *trackende* landen is de band tussen sociale afkomst en ongekwalificeerd uitstromen groter. Binnenkort zullen we ook meer zicht kunnen krijgen op de impact van vroege *tracking* op de vaardigheden van volwassenen, wanneer de uitkomsten van PIAAC (een internationaal vergelijkende test naar de lees- en wiskunde vaardigheden van de volwassenenpopulatie) geanalyseerd zullen kunnen worden.

2.3 Conclusie

De wetenschappelijke literatuur toont duidelijk aan dat *vroege tracking* het effect van SES op leerprestaties vergroot. Een hele batterij aan studies – met uiteenlopende onderzoekdesigns en op basis van verschillende datasets – heeft dit bevestigd. Op basis van empirische gegevens kan men dan ook enkel besluiten dat *een latere oriëntering de rechtvaardigheid van het onderwijs ten goede zal komen*.

Van alle in dit hoofdstuk genoemde studies – gepubliceerd in peer-reviewede wetenschappelijke tijdschriften - is helaas geen spoor terug te vinden in het rapport van Duyck en Anseel. Hierdoor hebben ze een kans gemist om het onderwijsdebat te objectiveren met wetenschappelijk onderbouwde uitspraken.

3 | Vroege tracking en kwaliteit (cognitieve effecten)

Dit hoofdstuk gaat dieper in op wat het internationaal vergelijkend onderzoek ons leert over de invloed van *vroege tracking* op de *kwaliteit* van het onderwijssysteem, d.w.z. het algemene niveau van de leerprestaties (cognitieve effecten).

3.1 De bewering van Duyck en Anseel

Duyck en Anseel claimen dat er

“overtuigende wetenschappelijke evidentie [is] dat vroege tracking een positief effect heeft op de leerprestaties”.

Hun argumentatie hiervoor valt uiteen in twee delen.

Ten eerste verwijzen ze naar één landenvergelijkende studie waarin zou worden aangetoond dat *vroege tracking* een positief effect heeft op de leerprestaties (Rindermann & Ceci (2009)). Dit is een eerder selectieve lezing van het onderzoek. Tientallen studies hebben immers net aangetoond dat *vroege tracking* géén positieve cognitieve effecten heeft. Al die studies ontbreken echter in de literatuurstudie van Duyck en Anseel. In sectie 3.2.1 belichten we deze uitgebreide wetenschappelijke literatuur, en leggen we ook uit hoe we die éne afwijkende conclusie van Rindermann en Ceci (2009) kunnen begrijpen.

Daarnaast baseren Duyck en Anseel hun bewering op een verzameling studies die positieve effecten laten zien van het tijdelijk en flexibel groeperen van leerlingen naar niveau: Kulik & Kulik (1992), Reis & Renzulli (2010), en Tieso (2002). Deze studies onderzoeken echter groeperingsvormen die helemaal niets te maken hebben met *early tracking*. In de door de studies onderzochte praktijken werden leerlingen uit het basisonderwijs één of enkele uren per week in een niveaugroepje ingedeeld (waarbij van niveaugroep kon worden gewisseld als dat nodig bleek). Dit soort groeperingsvormen komt ongeveer overeen met de niveau-leesgroepjes die we kennen uit de Vlaamse basisscholen. Alleszins gaat het hier duidelijk om iets helemaal anders dan *tracking*, waarbij leerlingen voor hun volledige schoolloopbaan in sterk van elkaar gescheiden onderwijsvormen worden geplaatst, elk met hun eigen invulling en imago. In de betrokken studies wordt er dan ook expliciet voor gewaarschuwd om de onderzoeksresultaten – over de effecten van niveaugroepjes – niet te gebruiken in een discussie over *tracking*. De foute interpretatie door Duyck en Anseel van deze studies, tégen die aanwijzingen in, wordt meer in detail behandeld in sectie 3.2.2.

3.2 Wat het onderwijskundig onderzoek ons leert

3.2.1 Landenvergelijkend onderzoek

In deze paragraaf geven we een overzicht van de resultaten van het landenvergelijkend onderzoek naar de effecten van *vroege tracking* op leerprestaties. Zoals gezegd blijkt de literatuurstudie van Duyck en Anseel hierover erg onvolledig: hun conclusie is gebaseerd op één landenvergelijkend onderzoek. Nochtans bestaat ook over dat onderwerp een veelheid aan wetenschappelijk materiaal.

Van de Werfhorst & Mijs (2010) vatten de algemene conclusies uit de empirische evidentie als volgt samen:

“The evidence with regard to efficiency in learning, operationalized by the average performance in a country, shows that educational differentiation [early tracking] leads to lower, rather than higher, average achievement in a number of subjects. (...) There is no evidence that average performance could be improved by allowing for higher inequality by means of educational differentiation (early tracking).”

Uit een objectieve wetenschappelijke review komt dus de boodschap naar voren dat *vroege tracking* géén positieve cognitieve effecten heeft. De betere presteerders winnen praktisch niets bij *vroege tracking*, terwijl de zwakkere presteerders er sterk op achteruitgaan.

Hieronder bespreken we opnieuw kort enkele van de studies die deze vaststelling empirisch onderbouwen.

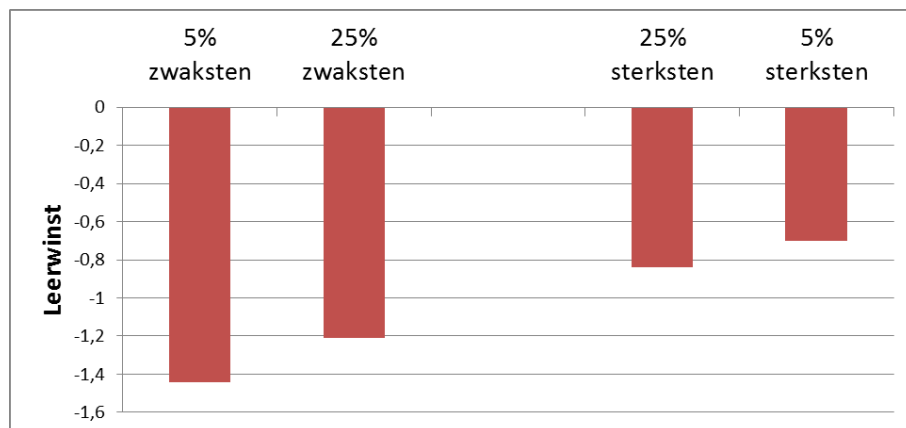
1. Een goede illustratie van de effecten van *vroege tracking* op cognitieve prestaties is de studie van Hanushek & Woessmann (2006). Zij maken gebruik van een *differences-in-differences*-design (zie sectie 1.4). Hierbij worden de prestaties van landen in het secundair onderwijs (PISA of TIMSS-2^e middelbaar) gecorrigeerd voor hun “startpositie” in het basisonderwijs (TIMSS-4^e leerjaar). Op die manier wordt de impact van versturende variabelen (bv. welvaart) geneutraliseerd: de versturende variabelen beïnvloeden zowel de leerprestaties in het basis als in het secundair onderwijs. Op het eerste meetpunt (4^{de} leerjaar) zijn er echter nog geen verschillen tussen landen qua *tracking*: het basisonderwijs is overal nog *comprehensief*. Op het tweede meetpunt (15 jaar) zijn die verschillen er wel: in een land als Duitsland hebben leerlingen dan al vijf jaar de effecten van *tracking* ondervonden, terwijl in *comprehensieve* systemen alle leerlingen nog samen zitten. Het effect van *tracking* kan dus worden afgeleid uit verschillen in de *toename* aan vaardigheden tussen de twee meetpunten. (In hun paper verwijzen Duyck en Anseel heel kort naar de studie van Hanushek en Woessmann. Ze beweren echter dat die niet geloofwaardig is omdat “*alle cross-nationale verschillen (andere dan vroege tracking) buiten beschouwing worden gelaten*”. Hun kritiek gaat dus volledig voorbij aan wat precies de opzet is van de *differences-in-differences*-techniek: door te controleren voor de startpositie worden de andere verschillen tussen landen uitgezuiverd, omdat die verschillen ook de startpositie moeten hebben beïnvloed.)

Hanushek en Woessmann bekijken acht combinaties van een meetpunt in het lager onderwijs (TIMSS of PIRLS, in het 4^e leerjaar) met een meetpunt in het secundair onderwijs (PISA op 15 jaar of TIMSS in het 2^e middelbaar). De meest interessante vergelijkingen zijn die waarin het tweede meetpunt PISA is, aangezien de populatie in PISA al 1 à 2 jaar ouder is dan die in TIMSS: als *tracking* werkelijk een effect heeft op de cognitieve effecten, dan verwachten we dat dat effect duidelijker wordt naarmate de “behandeling” (*tracking*) langer wordt toegepast.

De resultaten van deze vergelijkingen weerleggen de bewering dat *vroege tracking* positieve cognitieve effecten zou hebben. Voor de combinaties waarin PISA het tweede meetpunt vormt – en waarin die positieve effecten dus het meest zichtbaar zouden moeten zijn – is het effect van *vroege tracking* op de cognitieve prestaties zelfs altijd sterk significant *negatief* (!). Ook wanneer lees- en wiskundige vaardigheden worden getest met TIMSS (2^e middelbaar als eindpunt) als tweede meetpunt, is het effect van *vroege tracking* steeds negatief, zij het minder uitgesproken, aangezien de leerlingen dan nog maar relatief weinig tijd hebben doorgebracht in het *tracking*-regime. Wanneer natuurkundige kennis getest wordt (dit kan enkel met TIMSS, dus opnieuw met het 2^e middelbaar als eindpunt) is het effect dubbelzinnig en zou *vroege tracking* soms ook kleine positieve effecten kunnen hebben.

De analyse suggereert dus dat een vroege oriëntering van leerlingen géén positieve cognitieve effecten heeft. Het effect van vroege oriëntering is zelfs eerder negatief: voor de basisvaardigheden (wiskunde en leesvaardigheid) is het effect van *vroege tracking* in alle combinaties negatief, en vooral: het negatieve effect wordt sterker naarmate leerlingen ouder worden. Dat weerlegt volledig de hypothese dat *tracking* positieve cognitieve effecten zou hebben: als die er zouden zijn, zouden ze net beter zichtbaar moeten worden naarmate de *tracking* meer tijd heeft gehad om zijn werk te doen. We stellen echter net het omgekeerde vast.

Interessant is verder dat Hanushek en Woessmann niet alleen het *gemiddelde* niveau bekijken, maar dat ze de effecten van *vroege tracking* ook uitsplitsen tussen sterke en zwakke presteerders. Figuur 3 geeft voor de verschillende groepen leerlingen aan in welke mate *vroege tracking* hun prestaties beïnvloedt. “Prestaties” wordt hier opnieuw gekwantificeerd als de leerwinst die wordt geboekt tussen de twee meetpunten in. Voor de zwakke leerlingen is het effect van *vroege tracking* uiterst negatief: zwakke leerlingen hebben dus erg veel baat bij een uitgestelde studiekeuze. Toch halen ook de sterkere leerlingen geen voordeel uit *vroege tracking*. Ook zij presteren beter in de comprehensieve landen, al is het effect voor hen minder groot. De zwakke leerlingen zijn dus de dupe van *vroege tracking*, zonder dat dit duidelijke voordelen heeft voor de sterke leerlingen.



Figuur 3: Negatieve effecten van vroege tracking op leerwinst (in z-score) voor verschillende groepen leerlingen, overgenomen uit Hanushek & Woessmann (2006)

2. Het onderzoek van Hanushek en Woessmann vertrok van een differences-in-differences benadering, waarin prestaties in het secundair worden gecontroleerd voor prestaties in het basisonderwijs. Andere onderzoeken kozen voor de rechtstreekse controle voor verschillen tussen landen door opname in het model (zie 1.4). Ook met deze techniek is de conclusie echter gelijkloidend: *vroege tracking* heeft geen positieve cognitieve effecten, wel integendeel. Zo besluit Horn (2009) op basis van een multilevelanalyse waarin expliciet wordt gecontroleerd voor andere

kenmerken van het schoolsysteem dat *vroege tracking* geen positieve effecten heeft op de cognitieve prestaties van leerlingen:

“The idea that stratification [early tracking] overall enhances student performance can be rejected. Early age of selection (...) correlates slightly with lower effectiveness.”

3. Ook het landenvergelijkend onderzoek van Woessmann, Luedemann, Schuetz, and West (2009) bevestigt deze vaststelling: *vroege tracking* heeft géén positieve cognitieve effecten:

Early tracking does not seem to influence average student achievement in OECD countries. The results are inconsistent with the claims of supporters of tracking, yet provide strong additional support for the concerns voiced by critics.

4. Controlerend voor de economische context vonden ook Duru-Bellat & Suchaut (2005) dat *vroege tracking* geen positieve effecten heeft, niet op het gemiddelde niveau maar ook niet op het niveau van de beste presteerders. Meer zelfs, in comprehensieve landen doen ook de beste presteerders het zelfs nog iets beter dan bij de *vroege trackers*. Ook voor de subgroep van excellente studenten heeft *vroege tracking* dus geen uitgesproken voordelen:

“(...) Students in countries with earliest streaming perform slightly less well on average (...) The proportion of excellent students appears lower in countries with early streaming.”

5. Ook andere datasets geven steun aan de vaststelling dat *vroege tracking* niet nodig is om tot kwaliteitsvol onderwijs te komen. Zo besluiten Schütz, Ursprung & Woessmann (2008) op basis van TIMSS (2^e middelbaar):

“tracking does not exert a statistically significant direct effect on student performance.”

6. Door landenvergelijkende onderzoek te repliceren op deelstaten binnen één federaal land, waartussen de socio-economische en culturele verschillen relatief beperkt zijn, kan men verder hard maken dat de hierboven gerapporteerde effecten niet verstoord werden door andere verschillen tussen landen (zie 1.4). Woessmann (2010) laat zo opnieuw zien dat ook binnen Duitsland een keuze voor *vroege tracking* geen positief effect heeft op leerprestaties:

“A lot of public discussion surrounds the worry that there may be a tradeoff between achieving efficiency and equity in educational outcomes. [However], institutional measures associated with the slope of the socio-economic gradient, namely age of first tracking and number of tracks, are unrelated to mean performance. Thus, the performance level does not have to be traded off against equity.”

7. We kunnen ook opnieuw het effect van *tracking* gaan beoordelen op basis van quasi-experimentele situaties, waarbij de effecten van onderwijshervormingen zélf worden bekeken. Daalden de prestaties van leerlingen in landen waar de studieoriëntering werd opgeschoven?

Ook hier is de boodschap duidelijk: neen, het uitstellen van de oriëntering gaat niet ten koste van de kwaliteit. Pekkarinen, Uusitalo, and Pekkala Kerr (2013) laten bijvoorbeeld zien dat de vaardigheden

van jonge Finnen erop vooruit gingen nadat het Finse onderwijs in de jaren '70 werd hervormd (oriëntering op 16 jaar i.p.v. 11 jaar). Kansarme kinderen gingen er massaal op vooruit, maar ook de beter gesitueerden bleken niets te verliezen bij een latere oriëntering. Het effect was bovendien het grootst in die studiegebieden waar de hervorming het meest doortastend werd doorgevoerd (leesvaardigheid). Voor wiskunde (waarvoor tot in 1985 nog met verschillende beheersingsniveaus werd gewerkt, wegens de aanvankelijke weerstand tegen de hervormingsplannen) bleek het effect kleiner (niet-significant).

Ook hier is dus geen spoor te vinden van vermeende *negatieve* effecten van de latere studieoriëntering. De weerstand die aanvankelijk ook in Finland bestond tegen de hervorming van het onderwijs, precies omwille van de vrees voor kwaliteitsverlies, werd dus weerlegd door de feiten:

“The comprehensive school reform faced intensive resistance. Most common arguments against the reform were that abolishing tracking would reduce the quality of education. [However] the effect of the reform on those with less educated parents was positive and statistically significant in the verbal test [and positive but insignificant in the other two tests]. The reform increased the average score for those with less educated parents by about 0.2 and [produced] a zero effect for men with highly educated parents.”

De goede prestaties van Finland, met name voor de zwakkere leerlingen maar *zonder* dat de sterkere leerlingen daarvan de dupe zijn, hebben dus wel degelijk iets te maken met de vormgeving van het Finse onderwijs (late tracking).

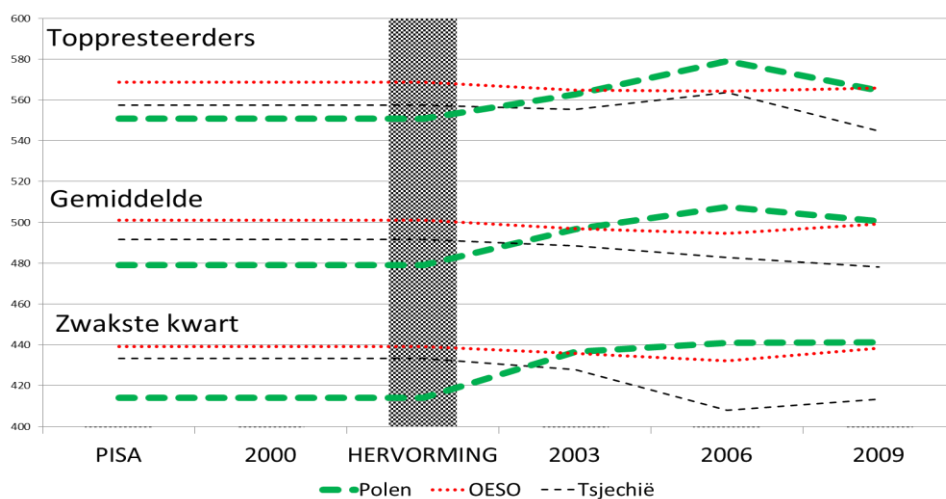
8. Een ander interessant quasi-experiment is de onderwijshervorming in Polen van 1999, waarbij de oriëntering 1 jaar werd uitgesteld. Jakubowski (2010) laat zien hoe Polen door deze hervorming beter is gaan presteren op de PISA-rankings:

“We find that that the reform was associated with significant improvements. Improvements in student outcomes in Poland measured by PISA have been impressive. In fact, in the first assessment [=before the reform] Poland ranked below the OECD country average in reading. In 2003, Poland reached the OECD average. By 2006, Poland scored above average, ranking 9th among all countries in the world.

Ook hier kunnen we trouwens de opdeling maken tussen effecten op zwakke en op sterke presteerders. Jakubowski laat zien dat de voordelen het grootst waren voor de leerlingen die anders in de “laagste” track zouden zijn terechtgekomen:

Performance improvement for potential students of former vocational schools is simulated to be higher than 100 points from 2000 to 2003 and 120 points from 2000 to 2006. This is more than one standard deviation of PISA scores in OECD countries, which is a dramatic improvement, hardly comparable to effects of any known educational policy. Obviously, these estimates are statistically significant, supporting the hypothesis that 15-year-old students who without the reform would be placed in vocational tracks heavily benefited from the reform.

Even belangrijk is dat ook de sterkste presteerders niet leden onder de hervorming. Dit kan bijvoorbeeld worden afgelezen aan de evolutie van de verschillende kwantielen in PISA. Figuur 4 laat de evolutie zien van de prestaties van de zwakke, gemiddelde en excellente scholieren in Polen, Tsjechië - een qua context met Polen vergelijkbaar land dat echter nog steeds vasthoudt aan een vroege tracking op 11 jaar - en het gemiddelde in OESO. Vóór de hervorming had Polen een serieuze achterstand zowel ten opzichte van Tsjechië als ten opzichte van de andere OESO-landen (van meer dan 20 punten). Dankzij de hervorming is dit gat echter volledig dichtgereden (zelfs als een recente neerwaartse knik mee in rekening wordt gebracht): Polen zit ondertussen al op of boven het OESO-gemiddelde, terwijl het Tsjechië al ver achter zich laat (met meer dan 20 punten). De bijzonder snelle inhaalbeweging is het duidelijkst voor het zwakste kwart, maar ze is zeker niet ten koste gegaan van de toppers.



Figuur 4: Het effect van latere oriëntering in Polen (na 2000) op de leerprestaties

Opnieuw valt op dat de hierboven geciteerde studies heel diverse onderzoeksmethodes gebruiken: het controleren voor verstorende variabelen, het controleren voor de startpositie (differences-in-differences), en het bekijken van concrete hervormingen binnen hetzelfde land (Finland 1973, Polen 1999). In al hun verscheidenheid hebben deze studies echter één ding gemeen: ze slagen er nooit in om *positieve* effecten van *vroege tracking* vast te stellen.

Maar hoe zit het dan met die ene studie die volgens Duyck en Anseel zou aantonen dat *vroege tracking* wél positieve cognitieve effecten heeft? De studie in kwestie (Rindermann & Ceci (2009)) is (net als een aantal hoger geciteerde studies) landenvergelijkend van opzet, waarbij men probeert verschillen tussen landen op te vangen met controlevariabelen (zie sectie 1.4). Rindermann en Ceci maken bovendien gebruik van dezelfde soorten gegevens (prestaties van landen op internationale

scholierentests) als heel wat hoger besproken onderzoeken. Hoe kan het dan dat hun onderzoek leidt tot resultaten die zo volkomen ingaan tegen de rest van de literatuur?

Een eerste belangrijke opmerking is dat Rindermann en Ceci de “cognitieve prestatie” van een land meten door een *optelsom* te maken van de gemiddelde scores van het land op alle scholierentests in hun databank. Dit lijkt misschien een goed idee, maar het is het niet. Onder die scholierentests zitten er immers ook een aantal die zijn afgenomen in het 4^e leerjaar! Het is uiteraard niet logisch om de effecten van *tracking* in het *secundair* onderwijs te willen aflezen aan de prestaties in het *lager* onderwijs⁴.

Een tweede opmerking is dat de door Rindermann en Ceci geconstrueerde dataset bijzonder heterogeen is. De meeste onderzoeken waar we hoger naar verwezen focusten zich op min of meer met elkaar vergelijkbare Westerse landen. Rindermann en Ceci breiden het studiegebied (willekeurig) uit met een aantal meer “exotische” landen zoals India, Iran en Zuid-Afrika. Op zich is dat geen slecht idee (meer datapunten). Maar aan het gebruik van zo’n diverse dataset hangt wel een heel belangrijke voorwaarde vast: het is uitermate belangrijk dat er doorgedreven wordt gecontroleerd voor alle verschillen tussen landen die niets met *tracking* te maken hebben (cf. sectie 1.4). Hoe diverser de landen in de dataset immers worden, hoe groter de kans dat een relevant verschil over het hoofd wordt gezien, en hoe sterker de schattingen er door verstoord kunnen worden.

En dit is precies wat er in de analyse van Rindermann en Ceci is misgelopen. Hoewel ze controleren voor een beperkt aantal variabelen, was deze controle duidelijk niet voldoende om de enorme verschillen tussen landen te neutraliseren. Een duidelijk voorbeeld is Zuid-Afrika, dat – ook na de controle van Rindermann en Ceci - heel slechte prestaties haalt. Dit heeft uiteraard alles te maken met de erfenis van de Apartheid – en niet met het feit dat Zuid-Afrika toevallig laat *trackt*. Een eenvoudige oefening is dan om de analyse van Rindermann en Ceci eens te herhalen voor een groep van landen waartussen de verschillen in sociaaleconomische en culturele context minder groot zijn, zoals de 34 OESO-landen. Als het effect reëel is, dan zouden we verwachten dat het ook binnen deze (grote) subset duidelijk wordt. Maar wat blijkt nu uit zo’n analyse (*in press*; voor een samenvatting zie www.indemarge.be)? Binnen de OESO kan er helemaal *geen positief effect van vroege tracking* worden vastgesteld. Het afwijkende resultaat van Rindermann en Ceci was dus volledig te wijten aan het feit dat ze een aantal ontwikkelingslanden aan de analyse toevoegden, dat ze onvoldoende corrigeerden voor de toegenomen diversiteit die hiervan het gevolg was, en dat dit de schatting van het effect van *vroege tracking* sterk verstoort. Voor een hoogontwikkelde regio als Vlaanderen is dat uiteraard geen relevante vergelijking.

⁴ Ter verdediging van Rindermann en Ceci merken we op dat hun artikel niet specifiek over het effect van *tracking* ging, maar dat ze in hun artikel verschillende onderwijspraktijken evalueren (zij het in afzonderlijke analyses, dus zonder bij de studie van één kenmerk te controleren voor de effecten van een ander). Ze gaan bv. ook na hoe prestaties afhangen van de mate waarin een land in onderwijs investeert. Voor dit soort verbanden – die niet specifiek over de structuur van het *secundair* onderwijs gaan – is het beter verdedigbaar om “prestaties” te definiëren als de gemiddelde score over het basis- en secundair onderwijs. Maar voor *tracking* is dat uiteraard niet zinvol.

3.2.2 Niet elke vorm van differentiëren is *tracking*

Het overzicht van landenvergelijkende studies naar de cognitieve effecten van *vroege tracking* levert dus een vrij duidelijk beeld op: er is geen bewijs te vinden voor positieve effecten van *vroege tracking*. Duyck en Anseel mobiliseren echter nog een tweede argument: Amerikaans onderzoek heeft aangetoond dat het flexibel en tijdelijk groeperen van leerlingen positieve effecten kan hebben op leerprestaties. Duyck en Anseel maken zich sterk dat die positieve effecten zonder meer kunnen worden getransfereerd naar een discussie over *vroege tracking*.

Maar is dat wel zo? Laten we de studies waar Duyck en Anseel naar verwijzen (Kulik & Kulik (1992), Reis & Renzulli (2010), en Tieso (2002)) eens van nabij bekijken. Duyck en Anseel claimen dat uit de verschillende onderzochte groeperingen de praktijk die in die studies wordt aangeduid als *within class grouping* “de Vlaamse situatie het best benadert” (p. 16). De praktijk waar het hier over gaat, wordt in de originele artikels als volgt omschreven:

“A teacher forms ability groups within a single classroom and provides each group with instruction appropriate to its level of aptitude. This type of grouping has been used frequently for reading and arithmetic instruction in elementary schools⁵. The teacher usually presents a lesson to one of the subgroups while the remaining groups engage in other activities.” (Kulik & Kulik (1992))

“This practice groups students within the same class into smaller groups for specific activities and purposes. Typically, the teacher presents a lesson to the whole class and then places students into small groups based on demonstrated performance, interests, levels of prior knowledge, and the like. This arrangement allows students to move into and out of groups based on their current demonstrated achievement.” (Reis & Renzulli (2010))

Het mag meteen duidelijk zijn dat de oorspronkelijke auteurs iets heel anders hebben bestudeerd dan *tracking*, de opsplitsing in verschillende onderwijsvormen zoals we die in ons secundair onderwijs kennen⁶. We zetten de belangrijkste verschilpunten even op een rijtje.

- Bij *within class grouping* is de indeling *flexibel en tijdelijk*: afhankelijk van de prestaties worden leerlingen van groep gewisseld. Dit is iets heel anders dan de definitieve rigide opsplitsing in onderwijsvormen, waarin de leerling in principe voor de rest van zijn schoolcarrière blijft zitten.
- De indeling geldt bovendien enkel voor *één bepaald vak* en kan dus verschillen per vakdomein. Afhankelijk van het vak zit eenzelfde leerling dus in de sterke of de zwakke groep. Voor nog andere vakken zit de hele groep gewoon samen. Ook dit is onmogelijk te vergelijken met *tracking*, waarbij leerlingen voor hun hele schoolperiode in volledig van elkaar onderscheiden onderwijsvormen terecht komen, vaak in fysiek van elkaar gescheiden scholen.

Gegeven deze verschillen is het duidelijk dat we van *within class grouping* heel andere effecten mogen verwachten dan van *vroege tracking* (zie 1.2). Beide praktijken hebben gemeenschappelijk dat ze lesgeven efficiënter kunnen maken (specialisatie-hypothese). Bij *vroege tracking* wordt deze potentiële winst echter gehypothekeerd door typische problemen als misallocatie (het plaatsen van leerlingen

⁵ Van de 11 door Kulik en Kulik bestudeerde voorbeelden van *within class grouping* gaan er 8 over het basis en 3 over het secundair onderwijs. In 8 gevallen ging het om rekenlessen, in 1 om leesgroepjes, in 1 in schrijflessen, en in 1 geval om de drie samen. Het verschijnsel van niveauleesgroepjes in de basisschool, zoals we dat ook in Vlaanderen kennen, is dus een goede illustratie van deze manier van werken.

⁶ Om misverstanden te voorkomen: sommige Amerikaanse auteurs gebruiken de term *tracking* als verzamelterm voor alle vormen van groepeerpraktijken (de VS kent ons systeem van *vroege tracking* in algemeen en beroepsonderwijs immers helemaal niet). Deze semantische overeenkomst verhuult dus een manifest andere inhoud, zoals we hier aantonen.

in de verkeerde track, onder andere door de invloed van SES op studiekeuze) en negatieve omgevingseffecten (demotivatatie in de lagere tracks, ontstaan van minder op studeren gerichte schoolculturen, bijgestelde leerkracht verwachtingen – zie de in 1.2 geciteerde onderzoeken van Van Houtte). *Within class grouping* wordt daarentegen veel minder met die problemen geconfronteerd. Doordat de lesgroepen flexibel worden samengesteld, kunnen foute plaatsingen snel worden rechtgezet. Omdat leerlingen vlot van groepje kunnen wisselen, blijft “opklimmen” steeds een reële mogelijkheid. Doordat de groepen vak per vak anders zijn samengesteld (en de leerlingen doorgaans trouwens het leeuwendeel van de tijd gewoon samen zitten), hebben peer- en omgevingseffecten sowieso een veel kleinere impact. Last but not least is de bedoeling van niveaugroepen dat de zwakkere leerlingen intensiever begeleid worden, zodat ze de rest van de klas zo snel mogelijk kunnen bijbenen – iets wat tussen onderwijsvormen helemaal niet gebeurt.

Dit zijn hoegenaamd geen triviale verschillen. Dat wordt trouwens ook gewoon bevestigd in de oorspronkelijke studies zélf. Voor Reis & Renzulli (2010) is het flexibele en tijdelijke karakter van de door hen bestudeerde groeperingsvormen hét doorslaggevende ingrediënt van hun succes:

“The major advantage of flexible grouping is the temporary nature of the groups. Students are assessed frequently for growth and reassigned to different groups based on that assessment.”

Het is precies op dat punt waar *vroege tracking*, de rigide opdeling in aso/tso/bsso/kso, zich van de hier bestudeerde groeperingsvormen onderscheidt. Reis & Renzulli (2010) benadrukken dan ook uitdrukkelijk dat het effect van de door hun onderzochte groeperingsvormen in geen geval mag worden toegepast op *tracking*!

“Although research on tracking has been shown to produce detrimental effects for some students (Oakes, 2005), we make a distinction between tracking and instructional grouping. We define tracking as the permanent placement of students into a class that is often remedial or advanced in nature with little chance of exit or entrance over the years. In contrast, several types of instructional grouping exist for academically talented students, and the ones reviewed in this article enable flexible movement in and out of grouping patterns.”

Ook Tieso (2003), een andere door Duyck en Anseel aangehaalde bron, legt veel nadruk op het verschil tussen de door haar bestudeerde flexibele groeperingsvormen (*ability grouping*) en de rigide vorm van *tracking* in aso/tso/bsso/kso die ook bij haar weinig genade kent. In een artikel onder de niet mis te verstane titel “*Ability Grouping Is Not Just Tracking Anymore*” legt ze uit waarom de permanente indeling in onderwijsvormen “*waaruit men nooit ontsnapt*” onaanvaardbaar is geworden, en waarom flexibele tijdelijke groeperingsvormen daar een alternatief voor kunnen bieden:

“In the past, ability grouping has been equated with tracking, a permanent and now unacceptable approach in which students are assessed based on prior achievement or measured intelligence and placed into streams or tracks from which they never escape, a situation that has created problems for advocates of equity and equality. The present and future of ability grouping lies in the flexible use of grouping.”

De door Duyck en Anseel aangehaalde studies gaan dus over flexibele en tijdelijke differentiatievormen, waarvan het effect in geen geval kan worden gelijkgesteld met dat van *tracking*. Die studies kunnen dan ook niet zomaar worden geïnterpreteerd als een bewijs voor de gunstige cognitieve voordelen van *tracking*, zoals Duyck en Anseel doen. De studies zijn wél waardevol omdat ze aantonen dat een flexibele vorm van differentiatie haar nut heeft. Dit is een gegeven dat perfect binnen comprehensieve systemen kan worden ingepast: binnenklasdifferentiatie en andere vormen van flexibele *within-class grouping* zijn vaak een belangrijk onderdeel van die systemen. Zoals we hoger al beschreven, worden bijvoorbeeld Finse leerlingen die voor een bepaald vak moeilijkheden ondervinden intensief bijgewerkt zodat ze later weer kunnen aanpakken. Dit is

precies het soort tijdelijke en flexibele differentiatie dat positieve effecten kan generen – cf. de uitstekende scores van Finse leerlingen op de PISA-tests – zónder dat het leidt tot een rigide opdeling tussen onderwijsvormen (*tracking*).

3.3 Conclusie

En uitgebreide studie van de wetenschappelijke literatuur toont aan dat *vroege tracking* niet noodzakelijk is om tot kwaliteitsvol onderwijs te komen. Voor de sterke leerlingen biedt het geen meerwaarde, terwijl de zwakkere leerlingen lijden onder een te vroege opsplitsing. Deze vaststelling wordt onderbouwd door een veelheid aan studies. Ook de ervaring met onderwijshervormingen in verschillende landen toont aan dat een latere oriëntering niet hoeft te leiden tot een nivellering naar beneden, wel integendeel.

4 | Algemene conclusie en aanbevelingen

Uit een uitgebreide literatuurstudie blijkt dat het vroeg opsplitsen van leerlingen in onderwijsvormen een bedreiging vormt voor de leerprestaties van de zwakkere leerlingen, zonder dat het een echte meerwaarde vormt voor de goede presteerders. Daarnaast versterkt die vroege *tracking* het effect van sociale afkomst op prestaties, omdat kinderen met een lage socio-economische status vaak (onterecht) in minder goed presterende tracks terechtkomen. Beide conclusies liggen in lijn met de lessen die de OESO, de organisator van de PISA-testen, zelf uit het omvangrijke empirische bewijsmateriaal trekt (OECD (2012)).

Het rapport van Duyck en Anseel (2012), waarin net het tegenovergestelde wordt beweerd, vertoont bij nader inzien ernstige gebreken. Een hele stroom aan wetenschappelijke literatuur over de cognitieve en sociale effecten van vroege *tracking* (zie hoofdstukken 2 en 3) blijft in hun rapport onvermeld.

Op basis van de wetenschappelijke evidentie lijkt de voorgestelde hervorming van het secundair onderwijs, met zijn bredere eerste graad, dus wel degelijk een goed idee. Op tal van punten die essentieel zijn voor het welslagen van de hervorming geeft het Masterplan echter nog geen antwoord. Het wetenschappelijk onderzoek toont aan dat structuurhervormingen belangrijk zijn – de leeftijd van *tracking* heeft een effect – maar veel hangt af van de manier waarop de aangepaste structuur concreet zal worden ingevuld. Er bestaat weinig twijfel over het feit dat een vroege opsplitsing de kansen van zwakkere en sociaal achtergestelde leerlingen hypothekeert, maar énkél het uitstellen van de oriëntering zal niet voldoende zijn om hun achterstand ook effectief weg te werken. Een structuurhervorming is een *noodzakelijke* maar daarom nog geen *voldoende* voorwaarde voor succes. Een hervorming vereist ook een aangepaste omkadering, bv. de extra middelen voor bijkomende assistentie bij de remediëring. Daarnaast moeten ook de leerkrachten adequaat op hun (complexere) taken worden voorbereid, bv. bijkomende aandacht voor binnenklasdifferentiatie in de leerkrachtenopleiding en -bijscholing. Het Masterplan gaat nog niet in op die randvoorwaarden. Dit verklaart trouwens ook voor een deel waarom de onderwijsvakbonden erg terughoudend waren in hun analyse van de hervormingsplannen⁷.

Ondanks dit voorbehoud komt het *principe* achter de hervorming – het later oriënteren van leerlingen – uit het wetenschappelijk onderzoek naar voor als een waardevol element om tot een sociaal rechtvaardiger onderwijs te komen, zonder dat dit nadelig hoeft te zijn voor het algemene prestatieniveau.

⁷ Zie bv. de open brief van de onderwijsvakbond COC op <http://www.coc.be/files/newsitems/.613/20130602%20hervorming%20secundair%20onderwijs.pdf>.

5 | Referenties

Ammermüller, A. (2005), 'Educational Opportunities and the Role of Institutions', ZEW Discussion Papers 05-44.

Bol, Thijs & Van de Werfhorst, Herman (2013), 'Educational Systems and the Trade-off Between Labor Market Allocation and Equality of Educational Opportunity', *Comparative Education Review*, vol. 57, p. 285 - 308.

Boone, Simon & Van Houtte, Mieke (2012), 'Social inequalities in educational choice at the transition from primary to secondary education: a matter of rational calculation?', *Kultura i Edukacja / Culture and Education*, vol. 91, p. 188 - 214.

Brunello, Giorgio & Checchi, Daniele (2007), 'Does school tracking affect equality of opportunity? New international evidence', *Economic Policy*, vol. 22, p. 781 - 861.

Dupriez, Vincent & Dumay, Xavier (2006), 'Inequalities in school systems: effect of school structure or of society structure?', *Comparative Education*, vol. 42, p. 243 - 260.

Dupriez, Vincent, Dumay, Xavier, & Vause, Anne (2008), 'How Do School Systems Manage Pupils' Heterogeneity?', *Comparative Education Review*, vol. 52, p. 245 - 273.

Duru-Bellat, Marie & Suchaut, Bruno (2005), 'Organisation and Context, Efficiency and Equity of Educational Systems: what PISA tells us', *European Educational Research Journal*, vol. 4, p. 181 - 194.

Duyck, Wouter & Anseel, Frederik (2012), 'Gelijke Kansen, Gelijke Kinderen, Gelijke Klassen? Early Tracking in het Onderwijs', Itinera Institute Discussion Paper 2012/4.

Fischbein, Siv (1980). 'IQ and social class', *Intelligence*, vol. 4(1), p. 51-63.

Hanushek, E. A. & Woessmann, L. (2006), 'Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries', *Economic Journal*, vol. 116, p. C63 - C76.

Hanushek, Eric A. & Woessmann, Ludger (2010), 'The economics of international differences in educational achievement', National Bureau of Economic Research.

Horn, Daniel (2009), 'Age of selection counts: a cross-country analysis of educational institutions', *Educational research and evaluation*, vol. 15, p. 343 - 366.

Jakubowski, Maciej (2010), 'The Impact of the 1999 Education Reform in Poland', World Bank.

Kulik, James A. & Kulik, Chen Lin (1992), 'Meta-analytic findings on grouping programs', *Gifted Child Quarterly*, vol. 36, p. 73 - 77.

Lavrijsen, Jeroen & Nicaise, Ides (2013), Parental background and early school leaving, Steunpunt Studie- en Schoolloopbanen, Leuven

Metaforum KU Leuven (2012), 'Hervormingen in het secundair onderwijs', *Visietekst* 71 - 48.

OECD (2012), 'Equity and Quality in Education'.

Pekkarinen, Tuomas, Uusitalo, Roope, & Pekkala Kerr, Sari (2013), 'School Tracking and Development of Cognitive Skills', IZA Discussion Paper No. 4058.

Reis, Sally M. & Renzulli, Joseph S. (2010), 'Is there still a need for gifted education? An examination of current research', *Learning and individual differences*, vol. 20, p. 308 - 317.

Rindermann, Heiner & Ceci, Stephen J. (1-11-2009), 'Educational Policy and Country Outcomes in International Cognitive Competence Studies', *Perspectives on Psychological Science*, vol. 4, p. 551 - 568.

Schütz, Gabriela, Ursprung, Heinrich, & Woessmann, Ludger (1-5-2008), 'Education Policy and Equality of Opportunity', *Kyklos*, vol. 61, p. 279 - 308.

Tieso, Carol L. (2002), 'The Effects of Grouping and Curricular Practices on Intermediate Students' Math Achievement', *National Research Center on the Gifted and Talented*.

Tieso, Carol L. (2003), 'Ability grouping is not just tracking anymore', *Roeper Review*, vol. 26, p. 29 - 36.

Van Houtte, Mieke (2004), 'Tracking effects on school achievement: A quantitative explanation in terms of the academic culture of school staff', *American Journal of Education*, 110(4), 354-388.

Van Houtte, Mieke (2006), 'School type and academic culture: Evidence for the differentiation-polarization theory', *Journal of Curriculum Studies*, 38(3), 273-292.

Van Houtte, Mieke & Stevens, Peter (2010), 'The culture of futility and its impact on study culture in technical/vocational schools in Belgium', *Oxford Review of Education*, 36(1), 23-43

Van de Werfhorst, Herman & Mijs, Jonathan JB (2010), 'Achievement inequality and the institutional structure of educational systems: A comparative perspective', *Annual Review of Sociology*, vol. 36, p. 407 - 428.

Woessmann, Ludger, Luedemann, Elke, Schuetz, Gabriela, & West, M. (2009), 'School accountability, autonomy, and choice around the world', Edward Elgar, Cheltenham, 2009